

An evaluation framework for the Flemish integration policies

Montserrat González Garibay

Peter De Cuyper

Steunpunt Inburgering en Integratie



Steunpunt Inburgering en Integratie

Consortium Universiteit Antwerpen (CeMIS), Katholieke Universiteit Leuven (HIVA),
Universiteit Gent (SDL) en Universiteit Hasselt (SEIN)
Lange Nieuwstraat 55
B-2000 Antwerpen

www.steunpuntieni.be
steunpuntieni@ua.ac.be

© 2013 Steunpunt Inburgering en Integratie

D/2013/13180/4
NUR 747, 759
ISBN 9789082094305

CONTENTS

INTRODUCTION	3
1.OVERVIEW OF THE EVALUATION LITERATURE	7
1.1 WHAT IS AN EVALUATION?	7
1.2 WHAT IS AN EVALUATION FRAMEWORK?	8
1.2.1 Definition	8
1.2.2 The policy process	12
1.2.3 Evaluation criteria	16
1.2.4 The role of influence and control	21
1.2.5 The design of an evaluation framework	23
1.3. WHAT TYPES OF EVALUATIONS ARE THERE?	28
2. POLICY LESSONS	30
2.1 GENERAL LESSONS	30
2.2 LESSONS FOR FLEMISH RESEARCH ON INTEGRATION POLICIES	32
3. THE CONSTRUCTION OF A FLEMISH EVALUATION FRAMEWORK.....	34
3.1 INTRODUCTION	34
3.2 CONCEPTUAL REFLECTION AND SCOPE	36
3.2.1 Definition	36
3.2.2 Purposes of the evaluation.....	36
3.2.3 Scope of the evaluation.....	37
3.3 EVALUATION CRITERIA	38
3.3.1 Selection	38
3.3.2 Conceptualization	39
3.4 VIEW ON CAUSALITY	42
3.4.1 Criteria	42
3.4.2 Theory of change	43
4. CONCLUSION	47
REFERENCES	49

INTRODUCTION

During the last decades, European and OECD states have responded to increasing immigration flows by designing and implementing integration and civic integration policies. Flanders has not been an exception: in addition to already existing integration policies, it launched a civic integration policy in 2003¹.

In order to improve the quality of those policies, both the European and the Flemish level have emphasized policy learning, evidence-based policy-making, and policy evaluation in their policy discourse:

1. At the EU level, **Member States have agreed on a set of Common Basic Principles for immigrant integration policy** to guide their domestic responses to integration. In those principles, evaluation adopts a fundamental role: *"Developing clear goals, indicators and evaluation mechanisms are necessary to adjust policy, evaluate progress on integration and to make the exchange of information more effective"*². This makes clear that, even though the EU clearly states that the harmonization of immigration and integration policies is not one of its goals³, there is a large emphasis on the sharing of best practices and thus *a fortiori* on policy learning and policy transfer.
2. At the Flemish level, the **improvement of the effectiveness and efficiency (i.e. evaluation criteria)** of integration policies constitutes one of the policy challenges that shape current civic integration policies, and one of the principles guiding the current reform process of the integration sector as a whole⁴. The two concepts – effectiveness and efficiency – are two important criteria that refer to the *quality* of the integration policies. In other words, an effective and efficient policy is a successful policy.

¹ The difference between integration and civic integration policies should be noted here. Integration policies take place in the context of the mutual accommodation of newcomers and their receiving society. Civic integration policies (*inburgering*) group the instruments aimed specifically at enhancing newcomers' self-sufficiency, and include language courses, cultural orientation and labor market orientation.

² Council of the European Union, Document No. 14615/04 (Presse 321), 2004.

³ It should be noted that the contents of integration policies are by no means a community competence that might imply a potential harmonization across member states. EU-level policy making on integration takes place without binding provisions, by applying the Open Method of Coordination in which the main control mechanism is peer pressure. This is, however, not the case for immigration policies, where a certain harmonization has taken place through several directives. For an extended assessment of the European dimension of integration policies see Caviedes (2004) and Luedtke (2009).

⁴ Voorontwerp van decreet betreffende het Vlaamse integratie- en inburgeringsbeleid; Beleidsnota Inburgering & Integratie 2009-2014.

When we look at the way in which those principles are translated in practice, a fundamental problem emerges. There are no common standards to assess the quality of a policy, neither at the EU nor at the Flemish level:

1. In order to engage into policy transfer and policy learning at the EU level, **policies should be to some extent measured by comparable standards** in order to assess both their success and their transferability to other settings. Even though the success and failure of policies are often compared at the EU level by means of benchmarking indicators and qualitative assessments of best practices, it is not clear what successful policies are, given the variety of ways to define integration: whereas one country may define a successful policy in terms of employment, other countries may emphasize migrant integration in the education system.
2. Efficiency and effectiveness have not been defined by Flemish integration policies⁵.

In other words, both the EU and the Flemish level lack a framework for evaluating the success or failure of civic integration and integration policies. Such a framework is, however, relevant:

1. **Evaluations fulfill an accountability function:** they provide a basis for Parliaments and citizens to sanction or reward those in charge of integration policies for their performance. For the Flemish civic integration policies, Art. 26 of the Civic Integration Decree mandates the evaluation of policies every three years. That evaluation should be submitted to the Flemish Parliament.
2. **A shared definition of what is policy success lays the basis of a common discourse or “language” among policy actors.** This enables them to compare policies both across temporary and spatial settings, and to learn from each other.
3. **An evaluation framework allows us to identify the conditions underlying policy success,** what enhances the likelihood of successful policy transfer. This is crucial both at the European level (transfer of policies across states) and in the Flemish context, where the local character of integration policies creates a variety of policy practices.

⁵ It should be noted that, from the perspective of the Flemish government as a whole, attention has been paid to policy evaluation in the form of Regulatory Impact Analysis, efficiency analyses from the Audit Office and the Finance Inspection and ex-ante tests regarding topics such as children’s rights or local government. Those evaluations have, however, an ad hoc character and often refer to ex ante processes as opposed to the ex post evaluation of policies already implemented. Moreover, they have been conducted in domains unrelated to integration and civic integration. Within the domain, some implicit definitions have been provided by policy documents, especially in the context of the current reform of the integration sector, but the two terms have not been the subject of clear definitions.

4. **An evaluation framework increases the evidence-based contents of a policy:** it provides objective criteria on the basis of which policy decisions can be taken. Those criteria are also a basis for replicability.
5. **The existence of clear standards to define policy success increases the potential visibility of policies at the international level,** given the fact that a standardized discourse makes it easier for messages to be conveyed to audiences⁶.

In this context, the present paper focuses on the question *how the success of integration and civic integration policies can be defined in a Flemish policy setting*. In order to answer the question, we develop an evaluation framework that contains several standards or criteria to assess integration and civic integration policies.

The framework's main goal is to provide a blueprint for the evaluation of the integration sector *as a whole* as opposed to the evaluation of concrete policies such as civic integration (language, social and cultural orientation, labor market orientation). Therefore, it does not set out any methodological elements, as the type of information available and the needed approach (quantitative, qualitative) may vary per instrument.

The evaluation framework is meant for two main groups of actors from the integration field:

1. Policy makers can use the framework as a blueprint for the conduction of policy evaluations. Such a blueprint allows to systematize the existing knowledge on the success of policies, fulfill government's accountability function vis-à-vis the parliament and enhance policies' evidence basis. Moreover, it provides a basis for setting-up and updating policy monitoring systems and tools.
2. For researchers, the evaluation framework constitutes a tool to shape a common ground on which to define policy-related research agendas, share results and ensure that their findings are used for the continuation of the research.

The framework is developed in three steps:

1. Chapter 1 defines some basic concepts regarding the evaluation of integration policies, with the aim of making the reader familiar with the evaluation terminology and mindset. On the basis of the scientific literature on evaluation, the chapter defines the terms evaluation and evaluation frameworks, links the

⁶ Even though Flemish integration policies have already received the attention of international organizations (see for instance OECD, 2008b), the attention paid from the international academic community remains limited.

distinct phases of the policy process to the evaluation concept, explains what evaluation criteria are and identifies the different types of evaluations. All the concepts are extensively illustrated with examples from existing evaluations of integration policies, both from Flanders and other regions/countries. 2.

2. Drawing on the review, Chapter 2 draws some lessons regarding the evaluation of civic integration policies. On the one hand it identifies the main differences and similarities across the evaluations of integration policies that have been conducted abroad. It does so from a conceptual and methodological perspective (i.e. what do the researchers understand as “evaluation” and “evaluation criteria”? How are the evaluations organized?). On the other hand it provides a critical assessment of the Flemish research on integration policies on the basis of those lessons. context.
3. Chapter 3 proposes an evaluation framework for the Flemish civic integration policies. The framework covers the purposes of the evaluation and its scope, as well as a first selection and conceptualization of five evaluation criteria, and a causal model. That framework may be used to bridge the existing gap in the Flemish evaluations and as a blueprint for future evaluation studies.

The evaluation framework is followed by a conclusion in which the further steps in the construction of an evaluation infrastructure for Flemish integration and civic integration policies are described.

1. OVERVIEW OF THE EVALUATION LITERATURE

This section provides an overview of the theoretical literature on policy evaluation and on the way in which evaluation frameworks are assessed in practice within the domain of integration. It starts by providing a definition of policy evaluation, and then moves to a short review of evaluation frameworks and four features of those frameworks: the policy process, the evaluation criteria, the role of influence and control and the design of an evaluation framework. A third section focuses on the existing types of evaluations.

1.1 What is an evaluation?

Policy evaluation applies the methods of social-scientific research to the performance or the effects of certain policy programs or projects (Rossi, Lipsey, & Freeman, 2004; Royse, Thyer, Padgett, & Logan, 2006; Swanborn, 1999; Wholey, Hatry, & Newcomer, 1994). Policy evaluation has been defined by several authors:

1. Scriven (1991, in Shaw, Greene, & Mark, 2006) states that "evaluation refers to the process of determining the merit, worth or value of something, or the product of that process.... The evaluation process normally involves some identification of relevant standards of merit, worth or value; some investigation of the performance of the evaluands on these standards; and some integration or synthesis of the results to achieve an overall evaluation or set of associated evaluations".
2. Rossi & Friedman (1985, p. 19 in Shaw, Greene & Mark, 2006, p. 7) define evaluation research as "the systematic application of social research procedures in assessing the conceptualization and design, implementation, and utility of social intervention programs. In other words, evaluation research involves the use of social research methodologies to judge and to improve the planning, monitoring, effectiveness, and efficiency of health, education, welfare, and other human service programs".
3. Fournier (2005, p. 140) argues that evaluation contains both empirical and normative components, i.e. a judgment about the value of something.
4. Swanborn (1999, p. 12) describes evaluation research as practice-oriented scientific research that consists of the set-up of interventions in social life, the guidance of those interventions throughout their implementation and especially the evaluation of their effects.

The definitions and typologies mentioned above display two common features:

- An evaluation implies a normative decision or analysis about a certain social fact or process.

- That normative decision is clearly embedded in a certain number of criteria, which Scriven terms "standards of merit", and which Rossi and Friedman specify further in terms of effectiveness and efficiency. The contents of those criteria are not fixed: they may change throughout time and place according to the policy context.

Evaluation should be distinguished from monitoring (Auer & Kruppe, 1996). Monitoring refers to the systematic mapping of policy processes and results by means of the systematic collection of qualitative and quantitative data. Monitoring provides us with an overview of the evolution of the elements of the policy processes, but it does not give us any explanation of those evolutions. Evaluation, by contrast, is the establishment of relationships between the different elements of the monitored policy processes, which are then used as levers to improve policies. The difference between monitoring and evaluation can be easily illustrated with an example. The EU's Indicators of Immigrant integration, which map the participation of immigrants in several domains by means of indicators such as activity and unemployment rates, educational achievement and income, are an example of a monitoring activity (Eurostat, 2011). A study in which those indicators are analyzed as the product of certain policies, such as language or integration tests, is an example of an evaluation.

It should be noted, however, that the line between evaluation and monitoring is thin: some evaluation criteria such as goal attainment (cf. *infra*) can also be considered monitoring activities, and the two activities may influence each other. For instance, the availability of monitoring data can influence the design of evaluations, or monitoring systems can be set up in function of evaluation goals⁷.

1.2 What is an evaluation framework?

1.2.1 Definition

The above implies that evaluations need to be carried out on the basis of explicit, *a priori* established criteria, which are defined in an evaluation plan (Owen & Rogers, 1999, p. 49) or evaluation framework (Bonin, Roberts, & Zimmerman, 2008). These criteria need to supersede the individual evaluator's "subjective perspective"⁸ regarding what constitutes "good" policies (Royse et al, 2006, p. 20). In other words, one of the

⁷ The term "evaluation" is often mistakenly used in order to designate monitoring or other types of activities.

⁸ For instance, Yanasmayan & Foblets (2012) analyze the impact of integration policies on the integration of immigrants in Flanders according to their own (implicit) definition of the integration concept.

fundamental intentions of an evaluation is to generate a *common language* or *common discourse* that can be used in order to judge a policy's success or failure. The existence of such a discourse is the basis for evidence-based policy-making.

There is no consensus among scientists or policy makers on what an evaluation plan or framework should look like. The concrete form it adopts for a particular evaluation often depends on empirical questions such as the level at which the framework is applied (international, national, local, cf. *infra*), the purpose of the evaluation, the policy priorities and the available sources of information. Nevertheless, it is clear from the definitions above that we need information on at least two issues in order to set up an evaluation framework. On the one hand, we need information about the policy that we are going to evaluate. On the other hand, we need to define the criteria that will be used in order to judge that policy. We offer an overview of both issues in terms of Wauters' (in De Cuyper, 2012) framework, pictured below. That framework overlaps largely with the OECD's evaluation framework for development assistance⁹. The framework defines the policy process in terms of five components: policy needs, policy goals, policy inputs, policy outputs and policy effects. The criteria for evaluation are conceptualized as relevance, efficiency, impact and effectiveness. Each of those components is described and illustrated with examples in the following section.

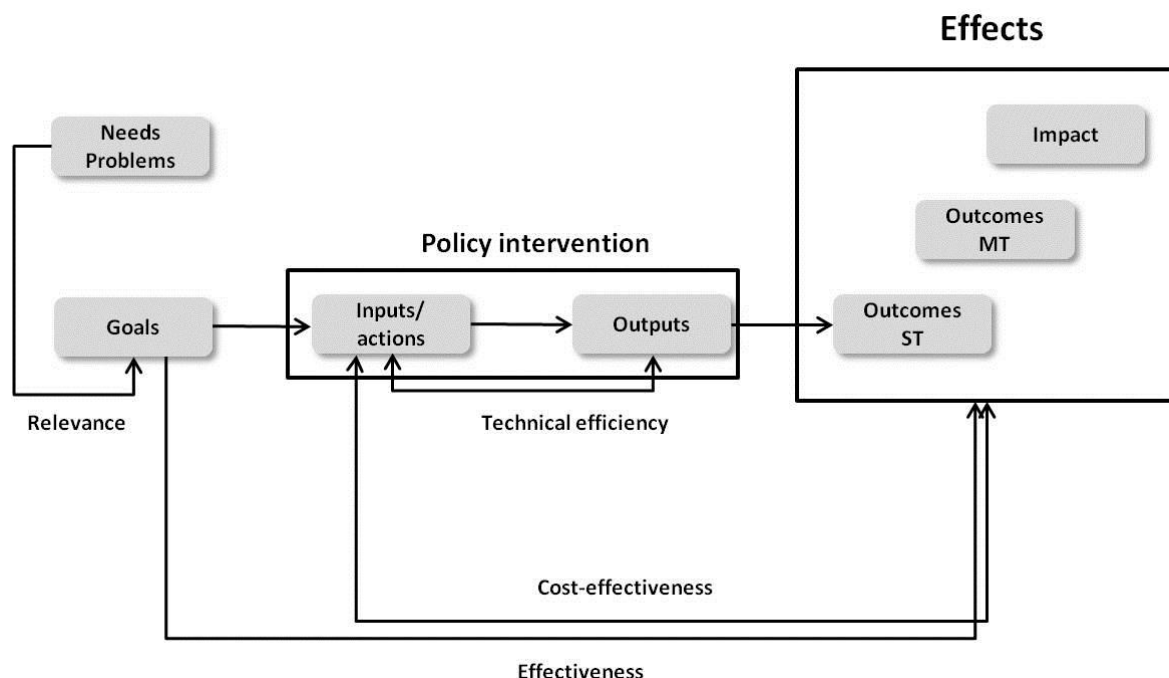


Figure 1. Wauters' evaluation framework
Source: Wauters (in De Cuyper, 2012)

⁹ In this context, we changed some of the terms used by Wauters with regard to policy's effects (cf. *infra*) in function of the OECD's criteria in order to not to confuse the reader.

In practice, we see that frameworks do not use all of the evaluation criteria purported above and, conversely, that they include other elements than the ones emphasized by Wauters, such as specifications about the agencies that conduct the evaluations, or deontological codes. The box below illustrates this variety with some examples from integration policies and other domains such as development cooperation.

Evaluation frameworks: policy examples

Evaluation frameworks can be found in most policy sectors (environment, education, health, development cooperation, integration) and belong to several levels (multilateral, national, regional, local) and types (domain-specific vs. instrument-specific). Moreover, they can be sector-specific or encompass several sectors. This box briefly sets out some examples, focusing on integration policies whenever possible.

a) Multilateral evaluation frameworks

Outside of the integration field, there are several multilateral evaluation frameworks. For example, we can find a myriad of multilateral frameworks within the development cooperation sector. The OECD issued in 2001 the Paris Declaration, which establishes five criteria along which development cooperation should be evaluated and a set of indicators to monitor those criteria (OECD, 2008a). Likewise, the United Nations Development Program provides some guiding principles, such as human development and managing for results; norms of ethics, impartiality, timeliness and quality; key concepts (evaluation, monitoring, outputs, outcomes), and roles and responsibilities for a number of policy actors such as UNDP administrators and evaluation officers¹⁰.

b) Integration in the European Union

By contrast, no common evaluation framework for integration policies has been developed at the multilateral or the EU level¹¹. Given the limited role of the European Union in integration policies, the activities of the European Commission are limited to facilitating contacts among member states for the exchange of information. However, the EU does look at whether policies are meeting their objectives based on the Member States' decision to monitor results based on harmonized data (Eurostat, 2011). There is, however, no attention paid to the way in which policies attain their goals or not.

c) National evaluation policies: South Africa

At the national level, both within and outside the European Union several countries have developed evaluation policies that are applied horizontally to all policy domains. That is the case of South Africa, where a monitoring and evaluation policy was developed. The policy contains information on the concept's monitoring and evaluation, system goals for the evaluations and a definition of the roles of several policy actors in the evaluation process (The Presidency, 2007).

d) National, domain-specific and instrument-specific policies and frameworks: Canada

In Canada, the evaluation of integration policies is nested in a larger national strategy: there is a national, horizontal-level Policy on Evaluation applied to all policy domains. On the basis of that policy, an integration evaluation policy was developed. That policy

¹⁰ See <http://web.undp.org/evaluation/policy.htm>

¹¹ It should also be noted that, except for Denmark, no frameworks for the evaluation of integration policies have been developed by any of the EU Member States.

forms, on its turn, the basis for the development of other evaluation policies for concrete policy instruments within the integration domain.

The national Policy on Evaluation was developed by the Treasury Board in 2009 (Treasury Board of Canada Secretariat, 2012). That policy is based on the concept “managing for results”, and regards evaluation as a management tool for policy-makers in order to improve the design of policies and programs. It also defines the evaluation policy’s objective, i.e. “to ensure that the government has timely, strategically focused, objective and evidence-based information on the performance of its policies, programs and initiatives to produce better results for Canadians”.

In addition, Canada’s Policy on Evaluation sets out the responsibilities of the different instances participating in the conduction of evaluations, and defines some evaluation standards with regard to planning, competency, integrity, measurement and analysis and reporting.

That policy forms the basis for the domain-specific Evaluation Policy developed by Citizen and Immigration Canada (CIC), which is the instance charged with integration policies. The CIC’s policy’s objective is “(...) to ensure that Citizenship and Immigration Canada (CIC) has an effective and independent evaluation function” (Citizenship and Immigration Canada, 2012). In other words, evaluations are a part of CIC’s tasks.

In addition to defining the roles and responsibilities of policy actors with regard to evaluation, the CIC framework defines some concepts (evaluation, performance...) and some evaluation standards: effectiveness, efficiency and relevance.

The CIC policy is, in its turn, applied by individual instruments’ evaluations, that set up specific evaluation frameworks. Those evaluations develop the standards above further. For instance, program relevance is divided into program need, program uniqueness and consistency with government priorities (Evaluation Division, 2010a).

d) Domain-specific frameworks: Denmark

Denmark developed a performance management system for the monitoring and evaluation of integration policies. In that framework, the policy process is conceptualized more or less in Wauter’s terms: the desired policy effect in the long term guide policy-making: they are used to define shorter-term outcomes, outputs that will lead to those outcomes, activities that will produce those outputs, and resources that will be deployed to conduct those activities (Moller Hansen, 2012).

A central element of the model is the theory of change, which was developed by government together with several stakeholders, and which sets out all of the elements of the policy process mentioned above and their mutual relationships. For a broader definition of theory of change, cf. *infra* (causality).

1.2.2 The policy process

As set out above, Wauters' model conceptualizes policies or, rather, the whole policy cycle (agenda-setting, policy design, policy implementation, cf. Howlett, Ramesh, & Perl, 2009) in terms of five elements: policy needs, policy goals, inputs, outputs and effects. We illustrate those elements below with examples from the integration literature.

Policy needs are situations, conditions or problems that trigger policies.

Policy needs: examples

In Ireland, the National Intercultural Health Strategy 2007-2012 was formulated on the basis of consultations with the target group (ethnic minorities) regarding their policy needs in the field of health. Those needs included, for instance, the improvement of language classes to tackle language barriers, the involvement of cultural mediators to deliver culturally sensitive services and accessible information (using by instance universal symbols). The needs were translated into policy recommendations and into policy goals (cf. infra) in the Strategy (Health Service Executive, 2008).

In Flanders, De Cuyper & Jacobs (2011) identify the policy needs regarding the provision of language courses in the context of insertion in the labor market. Among those needs they mention the lack of courses adapted to newcomers who are already working, and the need to provide integrated services to illiterate newcomers, for whom a language course is not enough as a step towards insertion in the labor market.

Policy goals are the product of or the answer to policy needs, and form the guidelines of policy interventions. Those goals can be situated at different levels: De Peuter, De Smedt, & Bouckaert (2007) divides them into strategic and operational. Strategic goals are situated at an abstract level. Operational goals are "intermediate" goals, of which the realization will contribute to the achievement of strategic goals. For example, enhancing the language proficiency of newcomers will contribute to their self-sufficiency. Due to their very nature, strategic goals are more difficult to measure than operational goals.

Policy goals: examples

The goals of the current Flemish civic integration policies are set out in the 2003 Civic Integration decree¹², and entail both strategic and operational components. Strategic goals are self-sufficiency, social cohesion, participation and shared citizenship. Operational goals are language proficiency and education level of newcomers, among other (cf. infra).

The development of policy goals can be an example of evidence-based policy making, as in the case of Norway, where the government introduced a set of goals in 2006. Those goals – which ultimately aim at ensuring that immigrants and their children achieve the same level of living conditions as the Norwegian-born populations – were established taking into account the available information, so that it would be possible to monitor their achievement and to evaluate them¹³ (IMDi & VOX, 2010).

Policy goals are not static and may change according to the political, economic and cultural context: integration policies in the Netherlands were based during previous decades on a cultural setting in which the collective dimension was important as a cultural horizon. This was reflected in the fact that multiculturalism, and later on assimilation, stood central as a goal of integration policies. Nowadays, the transition to a more individualized society is reflected in the integration policies' focus on "living together" (Verweij, 2012).

Policy inputs are the means invested to achieve the policy goals.

Policy inputs: examples

Policy inputs are not always specified. Denmark constitutes an exception in this regard: policy inputs for its integration policies as a whole are legislative competence, money (i.e. subsidies), campaigns, competent staff, leadership and knowledge. In this case, inputs are viewed from a static perspective as resources (Moller Hansen, 2012).

Policy inputs can also be described for a single instrument. In Ireland, the evaluation of a mentoring program in which newcomers are matched to locals who help them in their integration process describes inputs as "recruitment, training, induction, matching", etc. It regards inputs, in other words, from a dynamic perspective, in terms of actions or processes rather than resources (Healy, 2010).

In Flanders (De Cuyper & Wets, 2007), policy inputs have not been defined in policy documents. They have only been operationalized by evaluators in terms of the subsidies granted to finance civic integration policies.

¹² See http://ec.europa.eu/ewsi/UDRW/images/items/docl_1294_743451124.pdf

¹³ This constitutes an interesting illustration of the link between policy evaluation and policy monitoring: the evaluation design (i.e. the conceptualization of the goals to be evaluated).

Policy outputs are the direct results of the policy, and are realized by deploying the inputs.

Policy outputs: examples

Australia provides help to refugees through the Integrated Humanitarian Services Strategy (IHSS). Given the fact that that program is delivered by means of contracts with service providers, the outputs of the program are set in the service delivery contracts. The outputs are information provision, assessment/referral/short term counseling, training of other service providers and advice/consultancy to other service providers. The contracts also define units for each outputs (for instance, one unit for the first two outputs is a client serviced, and one unit for the third and fourth outputs is an hour of service) (Urbis, 2003).

In Flanders, the policy outputs have been assessed in both the 2007 and 2010 evaluations of civic integration policies. The outputs were defined in terms of the number of granted civic integration certificates to newcomers, the number of enrolments at the welcome offices and the number of civic integration contracts signed, (De Cuyper, Lamberts, & Pauwels, 2010; De Cuyper & Wets, 2007).

Policy effects are the consequences generated by the outputs at the broader societal level. They often imply a behavioral change of the policy's target group, and may or may not contribute to achieving the policy goals (desired vs. undesired effects). Effects are divided in three categories: short-term outcomes, mid-term outcomes and long-term impact. It is important in this context to stress the difference between outputs and effects. Outputs refer to the policy intervention's direct results, which can be traced back to the inputs and processes that fall more or less within the direct control of government. By contrast, effects are related to the broader policy results in terms of social processes taking place e the policy makers' control. Those processes are in first instance related to certain features of the target group (i.e. newcomers), such as language proficiency, knowledge of Flemish society and fit with the Flemish labor market. The impact should in principle coincide with the operational and strategic policy goals.

Policy effects: examples

According to the Danish evaluation and monitoring model, integration policies in general should lead to several outcomes: increased well-qualified immigration, increased employment and education, better language skills.... Examples of impact are better integration and social cohesion in Danish society¹⁴ (Moller Hansen, 2012).

In Germany and Australia, attention is paid by civic integration policies (language courses) to language proficiency as an effect. In addition, Australian programs also look at effects in terms of target group satisfaction, among other aspects. Portugal has also taken client satisfaction into account as a policy effect of its phone interpretation services (IOM, 2010; Schuller, Lochner, & Rother, 2011; Urbis, 2003).

The effects of integration policies are often assessed in terms of participation in the labor market. This is the case of Sweden, where short-term effects of integration policy have been defined as the participation of newcomers in non-subsidized employment (OECD, 2007).

In Flanders, no explicit or systematic classification of integration policies' effects has been conducted to date. However, the 2010 evaluation identified several dimensions along which effects of civic integration policies take place: labor, education, income, mindset, inter-ethnic contacts, social orientation, societal participation, health and housing (Pauwels & Lamberts, 2010).

¹⁴ It should be noted that the Danish classification is confusing if we look at its list of outputs, as there is no clear distinction between outputs and outcomes according to our definition (cf. supra): the list includes both matters that fall within the control of government (e.g. the flexibility of Danish language tuition) as matters that fall without the control of the government (e.g. resistance in society to radicalization and extremism). In other words, some of the Danish outputs should rather be classified as direct outcomes. For the whole list, see (Moller Hansen, 2012).

1.2.3 Evaluation criteria

The second component of the generic framework proposed by Wauters are the criteria by which the different elements of the policy processes are linked to each other, and which serve to judge the policy's performance¹⁵:

Policy relevance is defined as the extent to which the policy goals truly respond to policy needs. In the case of the integration policy we can apply this criterion to the relationship between policy goals and the needs of newcomer and the receiving society.

Policy relevance: examples

The Language Instruction for Newcomers to Canada (LINC) program looks at policy relevance in terms by assessing the need for language acquisition among newcomers. The 2010 evaluation found out on the basis of empirical information that "language was the most serious barrier newcomers faced to furthering their education or training and among the most serious barriers to finding employment". On these grounds, it concluded that the policy was relevant (Evaluation Division, 2010b).

In Flanders, no policy relevance assessments as such have been conducted to date. We can cite nevertheless two examples in which the relevance concept was assessed indirectly.

First, De Cuyper & Jacobs (2011) conducted an empirical assessment of Dutch language courses in which they assessed the extent to which the contents of the courses offered were adapted to the needs of newcomers who either work or are looking for work. The research concluded that offering courses outside working hours did respond to the needs of the population, and identified several ways to improve that policy to cover gaps in service provision. We can argue in this sense that the policy was implicitly judged to be relevant in general terms.

Second, the report on Social Impact (Wets, Seghers, Pauwels, De Cuyper, & Van Avermaet, 2012) concluded on the basis of empirical evidence that the introduction of "civic integration certifications"¹⁶ to be used in the process of applying for a job by participants to civil integration policies did not respond to the needs of employers, who rather wanted capable workers with whom to communicate in an acceptable manner. In other words, the policy's relevance from the employer's point of view was not confirmed.

Efficiency refers to the way in which policy inputs are combined in order to produce the policy results (outputs, effects). That relationship is divided into two sub-types. A first type of efficiency is cost-effectiveness, which points at the relationship between (short-term) direct outcomes of policies and the inputs which are invested in order to achieve those outputs. For instance, we can calculate how much it costs to provide a newcomer

¹⁵ The list is certainly not exhaustive. See below for some complementary examples.

¹⁶ Not to be confused with civic integration certificates. The certificate implies that the newcomer has followed the civic integration program, whereas the certification implies that he succeeded at the program's examinations.

with a job¹⁷. The second type of efficiency refers to the best possible ratio between inputs and outputs or, in other words, to the maximization of outputs for a given level of inputs. In other words, this second type of efficiency opens the black box of the processes of policy implementation in order to look at the way in which the inputs are deployed in order to produce certain outputs.

Efficiency: policy examples

In Canada, the evaluation of the Multiculturalism program, which allocates funds for the conduction of intercultural activities, included efficiency as a criterion. That efficiency was defined as the speed with which funds were allocated (Evaluation Division, 2012).

In a Swedish study on an experiment regarding job search guidance for newcomers, efficiency was defined as the added value in terms of income (for the newcomer who participated in the experiment and found a job), divided by the cost of the program for all those individuals who participated in the experiment and found a job.

In Flanders, both sub-types of efficiency have been assessed. On the one hand, cost-effectiveness was calculated by De Cuyper & Wets (2007) as the cost per signed civic integration contract (i.e. the cost per recruited person) and the cost per each civic integration program that was concluded. On the other hand, De Cuyper's (2010) analysis of policy processes contains an implicit reference to efficiency as the combination of inputs and outputs. It found, for instance, that a centralized system of referrals for language courses through welcome offices (onthaalbureaus) is more efficient for recruiting participants to social orientation courses than a system in which newcomers go directly to the centers charged with enrolments for Dutch courses, as this implies that several newcomers wanting to learn Dutch will not come into contact with social orientation courses.

Effectiveness is the extent to which an intervention realizes the policy goals for which it was created. In other words, it poses the question whether there is a causal link between the policy intervention and the policy results. Effectiveness can be sub-divided into two types, net and gross. Gross effectiveness refers to the measurement of a relationship between policy outputs and policy effects. Net effectiveness, on the other hand, attempts to establish a causal link between outputs and effects by excluding deadweight effects, i.e. those effects that would have taken place regardless of the policy intervention.

The distinction between gross and net effectiveness is easily illustrated by a fictional example. We measure the language proficiency of a group of newcomers before and after they have attended a Dutch course, and we compare it with the language proficiency of a similar group of newcomers who did not follow the course. We come to the conclusion that those newcomers who follow Dutch language courses have a higher language proficiency than those who do not follow the course. The difference between the language proficiency of those attending the course and those not attending the course

¹⁷ This definition of efficiency has already been applied in practice to integration policies by Scandinavian evaluations.

can be regarded as a measure of the gross effectiveness of Dutch language courses. We do not know, however, whether the difference between the language proficiency can be only ascribed to the language proficiency, or whether other factors – external courses, contacts with Flemish friends – have taken place. We do not know either whether some of the students would have acquired the same language proficiency by other means in absence of the policy intervention. Net effectiveness is, in other words, the *amount of effect that can be ascribed to the policy measure*.

Effectiveness is not a dichotomous concept. Gysen (2006) illustrates this point by locating effectiveness as one extreme of a continuum that goes from non-effectiveness to (full) effectiveness. If all policy goals are realized by a policy, the policy can be termed as effective. Conversely, if none of the goals is realized it can be termed as non-effective. If the goals are partially realized, the policy is located in the middle of the continuum and may be regarded as sub-effective. For instance, if the goal of a civic integration policy is to raise the proportion of newcomers at work by 10%, any rise lower than 10% that is ascribable to the policy (e.g. 7%) makes the policy subeffective. The policy can only be termed effective to the extent that it reaches the policy goal.

Effectiveness: policy examples

Effectiveness of integration policies is often assessed in terms of labor market participation (Germany, Norway, Sweden, Denmark) (OECD, 2007, 2012). One of the best-developed approaches is found in Denmark, where effectiveness is regarded as the comparison between the average duration of the period between the moment in which a newcomer obtains a residence permit and the beginning of his/her employment spell and the expected duration of that period. In order to compare the effectiveness of policies across municipalities while taking into account other factors (i.e. net effectiveness), the model takes into account additional variables such as the economic context (Moller Hansen, 2012).

In Australia, a program focusing on the provision of social orientation to refugees abroad defined effectiveness as client uptake, client satisfaction and client competences (Humanitarian Branch DIAC, 2009).

Client satisfaction is also a measure of effectiveness in Portuguese evaluations of the National Immigrant Support Centers. Effectiveness in this context is assessed by looking at the extent to which the centers help immigrants to overcome obstacles. This is measured from the client's perspective: the type of information obtained from the center, the timing of that information, the contribution of the centers to the immigrant's integration and the intention to use the center in the future are all measured (IOM, 2010).

As it has been mentioned, the above framework constitutes only one possible conceptualization of policy evaluations. In this context, there are other evaluation criteria than those highlighted above, such as:

Goal attainment (Swanborn, 1999, p. 59) refers to the extent to which goals are achieved regardless their cause, which may be located either within or outside policies.

For instance, we can measure the extent to which newcomers speak Dutch without necessarily knowing whether their language proficiency is to be ascribed to civic integration policies or to other factors, such as informal contacts with the Flemish population. In this sense, goal attainment can be related to gross effectiveness (cf. supra). Gysen (2006) argues that goal attainment is a necessary condition to attain full effectiveness.

Goal attainment: policy examples

The 2010 Flemish evaluation of civic integration policies looked at the extent of integration in the fields of language, labor, education, income, mindset, inter-ethnic contacts, social orientation, societal participation, health and housing, but it did not relate those outcomes with the policies in question. It did, in other words, measure the extent to which the government's goals were attained¹⁸ (Pauwels & Lamberts, 2010).

Policy coherence is the internal cohesion of a policy initiative within a certain policy field. Coherence is to be seen in the extent to which different measures contribute to the policy goals, and from the complementarity between policy goals (De Peuter et al., 2007, p. 115). For example, both language and social orientation courses contribute to newcomers' self-sufficiency.

Policy coherence: policy examples

Policy coherence is not often present in evaluations, as they tend to focus on a single policy instrument. However, in the Canadian Host program evaluation the uniqueness of the program vis-à-vis other policy tools is assessed. The assessment concludes that the program can be complementary to formal language instruction, since it provides newcomers with opportunities to improve language skills. There is, in other words, coherence between the two instruments as they both contribute to language learning (Evaluation Division, 2010a).

The Flemish policy program 2009-2014 for integration and civic integration attempts to enhance the coherence of the goals of three policy instruments (social orientation courses, language courses and career orientation courses) by means of an integrated goal framework. The actual coherence of the instruments has nevertheless not been assessed so far.

Consistency is defined as the extent to which positive and negative spillovers to other policy fields are respectively maximized and minimized. This points in other words at the coherence between different policy goals, and between those policy goals and overarching policy goals (De Peuter et al, 2007, p. 115). This criterion is named "appropriateness" by Owen & Rogers. An example of consistency is the overlap between the civic integration policies' goal of incorporating newcomers to the labor market and the labor market policy's goal of increasing the employment rate.

¹⁸ It should be noted that the research was termed "impact evaluation". We cannot speak of impact, however, as long as there is no causal connection between policy outputs and policy effects.

Consistency: policy examples

Canadian evaluations take into account the consistency of the measure being evaluated with the broader policy priorities of both the instance charged with integration policies (CIC, cf. infra) and the federal government. For instance, the evaluation of the HOST program looks at the way in which the program fits within Canadian federal legislation recognizing multiculturalism, and with the CIC's strategic outcomes, which focus on "the successful integration of newcomers into society and the promotion of Canadian citizenship through the implementation of integration programs".

Sustainability refers to the sustainable capacity of policies to tackle policy needs (Russon, 2005). The W.G. Kellogg Foundation (2004) emphasizes in this context the success of projects in developing "a strategy for the transition from short-term funding sources to long-term funding". This definition refers mainly to the sustainability of a program/policy (i.e. the financial capacity to extend the program). Conversely, the OECD's definition of sustainability looks at the sustainability of the *effects* rather than of the program itself (Development Assistance Committee, n.d.). It asks therefore the question "whether achievements are sustainable in the longer run".

Sustainability: policy examples

A German evaluation of integration courses defines sustainability as the emotional attachment to Germany, the language proficiency, the labor market insertion and the social contacts in German displayed by participants to the intervention one year after having finished the course (Schuller et al., 2011). It should be noted in this context that the German interpretation of sustainability overlaps with our definition of policy effects, rather than of the definition above.

1.2.4 The role of influence and control

The conceptualization provided above in terms of policy needs, goals, inputs, outputs and effects is only one way of looking at the policy process, which focuses essentially on the policy flow. We may choose instead to focus on other aspects of the policy process, as illustrated by the figure below.

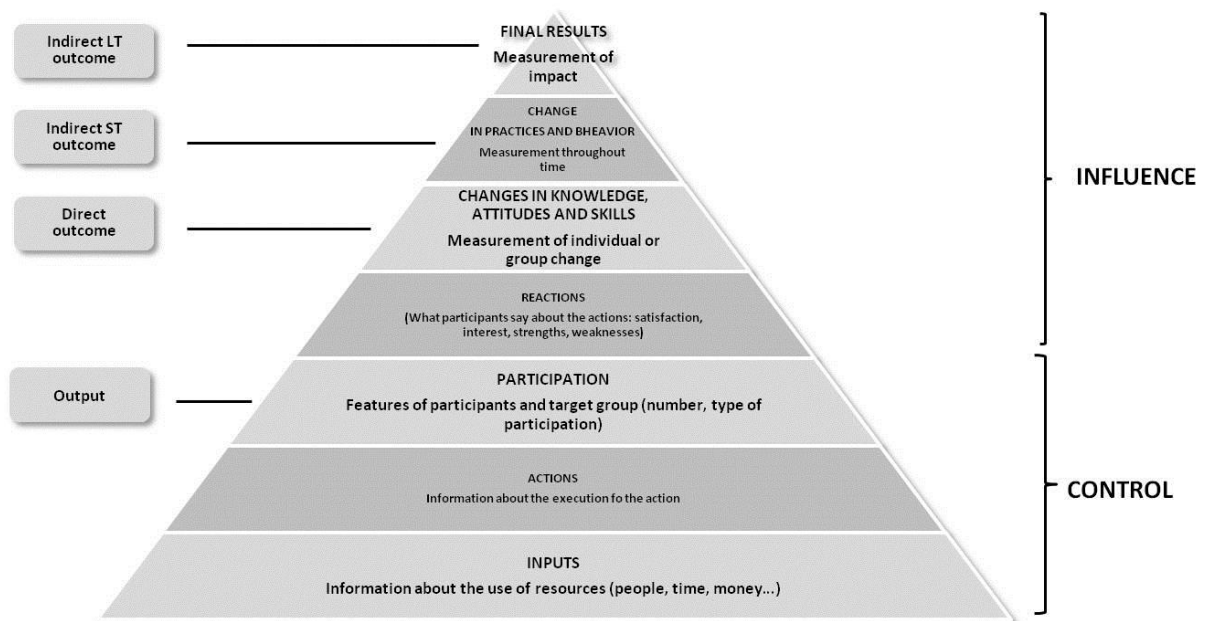


Figure 2. Influence and control in evaluation frameworks

The figure offers two complementary insights to the model of Figure 2 above. First, it incorporates the perspective of the target group, which is not taken into consideration by Wauters' model but which is nevertheless crucial to a policy's success, as policy interventions aim at behavioral change¹⁹ (cf. supra). In other words, it is important to take into consideration the opinions of the target group (newcomers) regarding the policy measures being evaluated.

¹⁹ It should be noted that the scope of the figure is limited to policy programs in which behavioral change is assumed to be the consequence of a change in attitudes and knowledge of the target group. Measures such as a change in immigration rules are therefore excluded.

The subjective perspective has been incorporated to the evaluation of integration policies in Canada: the evaluation of the Host program, which matches Canadians with newcomers to facilitate their integration, used surveys and focus groups of participants to the policies. The surveys and focus groups included questions about their perception of several aspects of the program, such as the matching procedure and the impacts in terms of adaptation to Canadian society (Evaluation Division, 2010a).

In Ireland, the impact of a similar program on mentoring was evaluated by means of interviews, in which the perceived effects of the policy were assessed.

In Flanders, the perspective of immigrants on integration policies has been integrated to a certain extent in two studies: Yanasmanayan and Foblets (2012) look at the implementation of civic integration policies through the eyes of their participants by means of interviews in order to identify their perceived effects in terms of integration.

Pauwels and Lamberts (2010) also used qualitative interviews in order to determine the effects of civic integration policies in Flanders: they interviewed newcomers on, among other, what they had learned during their courses and what they would change from those courses. A more recent report (Wets et al., 2012) looks at the way in which newcomers who participated in civic integration interventions look at the value that Flemish society attaches to their participation in those courses (civic effect).

Second, the figure makes clear that government's control over the policy results gradually diminishes along the chain of policy results. Whereas the last three segments of the pyramid (changes in knowledge and practices of the target group, final policy results) are the ultimate objective of the policy intervention, its possibilities to ensure change are limited. Government has direct control over policy inputs and over the actions by which those inputs are deployed, but only influence on those actions' effects.

1.2.5 The design of an evaluation framework

The way in which a concrete evaluation framework is designed depends strongly on the way in which causality is approached. When we look at the evaluation framework above, three issues regarding causality become clear. First, the establishment of causal connections is an essential component in the measurement of effectiveness, which is the most important indicator to determine the success or failure of a policy (i.e. has the policy triggered behavioral change in the desired direction?). Second, the fact that the ultimate policy goals are located outside government's sphere of control makes it difficult to establish causal relationships between policy inputs and the desired behavioral change. Third, causality is always present in the assessment of effectiveness. A policy may be partially or fully effective, but there is *always* a cause-effect relationship between the policy and its effects to speak of effectiveness. Methodologically, this implies that for a policy to be catalogued as effective we need to actually prove its causal link with the effects.

The causality issue is often approached in evaluation research as an **attribution problem**. Attribution refers to the relationship between the policy itself (inputs, actions by which inputs are deployed) and policy results (outputs and effects). Even if we come to the conclusion that the goals of a policy are attained, we do not know for sure whether they are the product or the results of the policy.

Some authors argue instead for the use of the term *contribution* instead of attribution. Contribution points at the role that a policy intervention has in bringing about certain outcomes *along with other factors*, instead of assuming that only the policy intervention is relevant for the policy results (Mayne, 2011). In other words, we attempt to look both at net effectiveness and at the other causal factors generating behavioral change.

The above can be illustrated by an example. We notice that an increase in the number of newcomers following Dutch language courses in municipality A during year X is accompanied by an increase in the employment rate of newcomers in the same municipality during year X+1. At first sight, we can conclude that following the course has led to a heightened labor market participation of newcomers.

However, if we take into account the economic context in our analysis, we notice that, due to the economic conjuncture, several jobs have been created in municipality A during year X+1. To determine the real contribution of the language courses to the heightened employment rate, we should compare the employment rate of those newcomers following the course with a group of newcomers who did not follow the course. The difference between the two groups is the policy's net-effectiveness.

The delicate role of causality in evaluation has a fourfold implication for evaluation research:

1. First, **it is necessary to develop strong theoretical frameworks** that set out the concrete mechanisms along which policies will achieve the desired effects and that, at the same time, integrate the context and the other possible factors playing a role.
2. Second, **a strict methodological approach is needed** in order to determine the policy's contribution to a certain outcome.
3. Third, **those evaluations which attempt to measure the aspects of policies characterized by government control, such as the extent to which outputs are produced, will provide more clear-cut conclusions** and certainty than those evaluations that assess those criteria situated outside the direct control sphere, such as effectiveness.
4. Fourth, the **establishment of clear causal connections is crucial to policy learning** and evidence-based policy making, as we need clear insights on what types of policy interventions work, and under which conditions.

From the above we retain the need of an adequate theoretical and methodological framework for evaluation to deal with the contribution problem in an adequate manner. There are several manners to do so in the evaluation framework. On the one hand, some authors such as Stern et al. (2012) have identified different types of causal relations, which are assessed by different methods:

Regularity. This term refers to tracing back the frequency with which causes and consequences appear together. Methodologically, this is often done on the basis of statistical techniques.

An example of regularity is the collection on data about newcomers that followed language courses in different municipalities, the job creation that took place in those municipalities and the employment rate of the newcomers, and the subsequent analysis of those data with statistical techniques such as multiple regressions.

Counterfactuals. They consist of comparing outcomes for populations that have participated in a policy intervention and groups that haven't (control groups). This can be done on the basis of experiments.

In the example above a counterfactual would imply a comparison, within a certain municipality, of the employment rate of those newcomers that followed language courses and those newcomers who didn't. The difference between the two groups that can be attributed to the policy is then the net effectiveness of policies. Such an approach has already been applied by Swedish and Danish evaluations, which compare participants to non-participants groups in order to determine the effects of civic integration policies on the labor market participation of newcomers.

Multiple causality. This approach attempts to trace back the possible combinations of causes that can lead to a certain outcome. This is often done by means of qualitative or configurational techniques such as Qualitative Comparative Analysis (QCA).

In the example above, multiple causality implies identifying the different paths that may lead to an enhanced employment rate of foreigners. That employment rate can either be the product of the economic conjuncture and the demand for certain skills that only the newcomers have, or alternatively a combination of Dutch language skills and a high number of vacancies that are not filled by the native Dutch-speaking population, such as positions in the care of the elderly or butcher.

Generative causality. This perspective focuses on causal mechanisms: we do not only look at *whether* policies are effective, but also at *how* that effectiveness is achieved. Therefore, it uses *theories of change* in which the mechanisms are depicted²⁰.

In our example of Dutch courses and employment, a generative vision of causality would imply that we depict the mechanisms according to which we think that the courses contribute to enhancing the newcomers' chances of employment. One of those mechanisms can be the fact that knowledge of Dutch allows newcomers to build a social network. In turn, they can make use of that network as an informal recruitment channel, so they are more successful at finding a job than those newcomers who do not speak Dutch. As a second step, we test those mechanisms to the available empirical information.

It should be noted that policy evaluations do not need to be limited to one of the above perspectives: depending on the strengths and weaknesses of each approach in combination with the specific policy issue at stake some characteristics of evaluation designs can be combined.

On the other hand, causality has been conceptualized in the field of social interventions in terms of an *effectiveness ladder* (Veerman & Van Yperen, 2007). The ladder is an instrument or model which allows to classify different types of evidence that evaluations may offer regarding the causal relationship between an intervention and its effects. It is built on the recognition of we can only have certainty of the effectiveness of a policy intervention by using strict methods (randomized control trials), but at the same time we can also use other types of evidence such as case studies or surveys in order to acquire some information on that effectiveness, albeit with less uncertainty. The model is summarized in the figure below.

²⁰ For a detailed example of a theory of change see De Cuyper, De Rick, & Gonzalez Garibay, 2012.

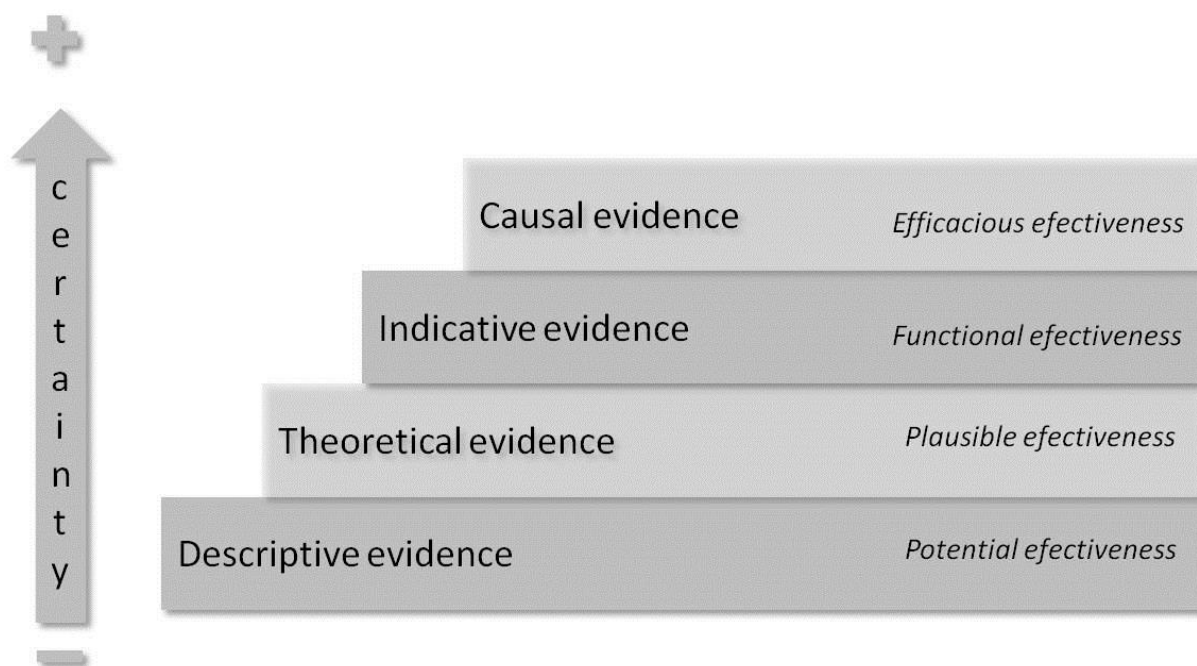


Figure 3. The effectiveness ladder

Source: inspired by Veerman & van Yperen (2007, p. 216).

As shown by the figure, the ladder is composed of four levels:

Descriptive. This type of evidence provides a description of the essential elements of the policy intervention such as goals, inputs and target groups. The evidence is obtained by methods such as descriptive and observational studies, document analysis or interviews. Even though they do not offer certainty about the effectiveness of a certain policy intervention, they provide us information about the *potential* effectiveness of interventions.

The Flemish evaluation of civic integration policies (De Cuyper et al., 2010) starts with an extensive description of integration policies and a rigorous analysis of the policy processes.

Theoretical. The second level in the ladder provides not only the description of a program, but also a theory of why the program should work and with whom, based on existing knowledge. This evidence is obtained by means of reviews, meta-analyses and expert knowledge studies. It allows us to say whether effectiveness is *plausible* or not.

Schibbel, Fazel, Robb, & Garner(2002) conducted a feasibility study for a systematic review of refugee integration policies. In that study, they looked for available evidence of the effectiveness of refugee integration policies in the scientific literature on the basis of clearly established criteria.

Indicative. This type of evidence demonstrates that the intervention leads to the desired outcomes (i.e. that it is effective). However, we do not know for sure which

aspect of the intervention caused the effects. The methods used to obtain this evidence are client satisfaction, goal attainment, monitoring, quality assurance, quasi-experimental and theory of change studies, as well as norm referenced approaches and benchmark studies. These methods allow to measure the *functional* effectiveness of policy interventions.

Several panel studies have been carried out in Scandinavia in order to explore the effectiveness of civic integration policies. For instance, Delander, Hammarstedt, Månsson, & Nyberg (2005) found on the basis of a quasi-experimental design that a workplace training program for difficult to place immigrants in Sweden was effective to ensure employment. De Cuyper e.a. (2010) found that newcomers who successfully complete a civic integration programme, participate more frequently in the labour market than newcomers who don't.

Causal. Causal evidence is built on Randomized Control Trials (RCT- and repeated case studies, which allow us to identify which parts of the intervention are responsible for which outcomes. It allows to measure the *efficacious* effectiveness of policy instruments, and offers the highest degree of certainty with regard to the effectiveness of the interventions.

Joona & Nekby (2012) found that an intensive counseling program with random assignment had a positive effect on immigrants' employment.

The two approaches to causality presented above offer complementary perspectives: whereas Stern's classification of causal relationships provides a theoretical conceptualization of causality, the effectiveness ladder does not describe the way in which the causal relationship occurs, but focuses rather on the type of evidence available and what that evidence tells us about effectiveness.

In this sense, the effectiveness ladder adds realism to Stern's classification, as it allows us to analyze the effectiveness of interventions even if we do not dispose of "perfect" evidence (i.e. from experiments) about the causal relationship between the policy and its effects. For instance, descriptive evidence can be used to assess the potential effectiveness of an intervention. Moreover, it orders the different types of evidence according to their causal strength.

1.3. What types of evaluations are there?

Evaluations can be classified along several dimensions. De Peuter et al (2007) provide an overview of several evaluation typologies according to distinct criteria:

Timing. Here they make a distinction between ex-ante, interim and ex-post evaluation. Ex-ante evaluations take place when a policy is being designed; interim evaluations are conducted when a policy is being implemented, and ex-post evaluations are carried out when the policy cycle has already been ended, i.e. after the policy implementation phase. This distinction partially overlaps with Swanborn's (1999) typology, which distinguishes plan evaluations (conducted after policy goals have been established), process-evaluations (which aim at adjusting policies during the implementation phase) and product evaluations, in which the results of the policy are assessed. This does not imply, however, that the categories cannot be combined. For instance, an assessment of effectiveness can take place both ex-ante and ex-post.

Contents vs. impact. Content evaluations focus on the structure, concept, processes and actions of a policy, whereas impact evaluations attempt to grasp the societal changes that a policy has brought about. In this sense, content evaluations are similar to process evaluations, whereas impact evaluations can be equated with Swanborn's product evaluations.

Evaluation criterion. Evaluations may focus on the different criteria set out above (efficiency, effectiveness, etc.).

Person conducting the evaluation. Internal evaluations are carried out by evaluators belonging to the instances carrying out or preparing the policy intervention, whereas external evaluations are conducted by independent evaluators. Both evaluation internal and external evaluations are however subjected to certain basic principles such as objectivity and neutrality.

Evaluation types: policy examples

In 2008, the Office of the Irish Minister of Integration “commissioned an independent review to assist in the development of a national English Language policy and framework for legally-resident adult immigrants. That review, composed of an analysis of foreign practices, written submissions and consultations and a survey of organizations providing language training to immigrants, can be considered to be such an ex ante evaluation, as it looked at the possible effectiveness of the proposed measure: it found out support for the development of the policy, as it would have positive long-term benefits for immigrants in terms of language skills, that help improve job opportunities; higher earnings, and educational opportunities (Horwath Consulting Ireland, Ramboll Management, & Matrix Knowledge Group, 2008).

The Flemish 2010 evaluation of civic integration policies is an example of a combination of a process and an ex post evaluation (De Cuyper, 2010; De Cuyper, Lamberts & Pauwels, 2010; De Cuyper, Lamberts, Pauwels & Vets, 2010; Pauwels & Lamberts 2010). On the one hand, it looked at the way in which policy processes unfolded: it tracked the implementation of civic integration processes from the first contact of newcomers with integration services until newcomers finished their program. On the other hand, it looked at the further consequences of policy in terms of goal attainment.

The Canadian Evaluation Policy focuses on internal evaluations, and sets out the roles and responsibilities of several actors within Canadian government instances (Deputy Heads, Departmental Heads of Evaluation, Departmental managers). By contrast, other countries such as Denmark, Australia and the Flemish region of Belgium conduct external evaluations (by research institutes in the case of Denmark and Flanders, and by consultancies in the case of Australia).

2. POLICY LESSONS

When looking at the definition of the evaluation framework and the examples provided throughout this section, we can identify several policy lessons, both at the general level and from a specific Flemish perspective.

2.1 General lessons

It is clear, from the overview, that evaluation frameworks and their application in concrete research are characterized both by plurality and by a number of common principles or characteristics:

1. There is a wide variety of evaluations of integration and civic integration policies:
 - a. ***Presence or absence of evaluation criteria.*** Evaluations can be based on evaluation frameworks and on clear criteria such as relevance, efficiency or effectiveness, such as in the Canadian and the Danish cases, or they can take place in an ad-hoc manner without thoroughly defining the criteria or the framework, as in the case of Yanasmayan and Foblets (2012). The advantages of a coherent evaluation framework and standards are clear: they entail a conceptual reflection in which the person or team conducting the evaluation is obliged to externalize his or her assumptions regarding the quality of policies and best practices. From the point of view of evidence-based policy-making, this is crucial: the construction of an evaluation framework makes us specify what we mean by policy success and failure.
 - b. ***Variety of evaluation frameworks.*** When evaluation frameworks are used, they can adopt several shapes. They can either focus on the contents of an evaluation (Denmark) or on the way in which the evaluation should be conducted (Canada). They can also include information on the policy process itself, as in the case of UNDP, or only specify the evaluation criteria that should be taken into consideration. In addition, they can specify few or several evaluation criteria.
 - c. ***Stand-alone or "embedded" evaluations.*** Evaluations can either be the product of stand-alone research, such as in the case of the Flemish evaluations, or be embedded in a larger evaluation system, such as the Canadian case, in which every evaluation (Host, LINC) is based on the CIC Evaluation Policy, and where the CIC Evaluation Policy itself is aligned with the Treasury Board's Policy on Evaluation. The embedded approach also has advantages: the fact that the Canadian evaluations are rooted in well-defined

principles allow for continuity and comparability throughout time. This facilitates policy-learning and thus evidence-based policy-making.

2. There are, however, a number of common elements to those evaluations that should not be overlooked:

- a. **Evaluation frameworks as a conceptual reflection.** In spite of the variety of the concrete way in which evaluation frameworks are shaped, they all include a *conceptual-reflective* component: they specify what is meant by an evaluation and the standards according to which that evaluation should happen. Those standards serve as a guiding principle for the conduction of the evaluation, as a standard against to which we can measure the quality of the evaluation and, as it has been stated above, a source of information for evidence-based policy-making.
- b. **Pervasive focus on causality.** Most evaluations attempt to assess the effectiveness of policies, whether that effectiveness is set in the future (e.g. Ireland) or in the past (e.g. Sweden, Denmark). At the same time, causality is approached in several ways: by qualitative or quantitative analysis, and by looking at long-term impact or short-term outcomes.
- c. **Selection of criteria.** Not all of the criteria assessed simultaneously. Most evaluations focus on one or two criteria, depending on the policy priorities, the available data and the characteristics of the measure.
- d. **Focus on labor market.** Most evaluations assessing the results of policies whether in terms of outputs or effects focus on labor market rather than on other policy domains such as cultural integration, income or education.

From the above lessons we can derive a few essential criteria that should be taken into account when developing an evaluation framework for the Flemish integration policies:

- An evaluation framework should make explicit what is meant by "evaluation" and define its own scope (all policies of a country, all policies within a certain domain, a specific policy instrument).
- An evaluation framework should specify the evaluation criteria to be taken into account.
- If the effectiveness of policies is among the selected criteria, the framework should specify a clear view of causality.

2.2 Lessons for Flemish research on integration policies

To date, several studies addressing civic integration and integration policies have been conducted in Flanders: they include the two evaluations mentioned above, the Yanasmayan and Foblets (2012) study on the newcomers' perceptions, and a few other studies (Verstraete A et al, 2001, Verstraete, Haertjens et al, 2001; Verstraete, Verbruggen & Cornelis 2000) on the potential shape for civic integration policies at the local level, a description of the target group, and the goals and contents of social orientation courses. When looking at the ensemble of studies in function of the lessons learned from the evaluations' overview provided above, we identify three main gaps in the research:

1. **No common evaluation framework.** There is no common evaluation framework or policy as in the case of Canada, and there is little communication between the different types of studies: the results of one research are not used as an input for the following assessment. For instance, the HIVA studies do not take into account the policy goals defined by Verstraete et al. There is, in other words, a need of *coordination* and *continuity* across studies.
2. **No conceptual reflection on policy process and evaluation criteria.** Evaluation elements and criteria are present in Flemish research, either in an explicit way, such as in the 2010 efficiency measurement, or in an implicit way, as in the case of policy relevance of language courses and social impact, or as in the case of outputs in the Yanasmayan & Foblets study. There is, in other words, a *lack of conceptual reflection* on the evaluation practice: we do not know why the authors selected a certain definition, and what the advantages of that definition are in the Flemish policy research context.
3. **No attention to causality.** Whereas the Danish and Swedish effectiveness measures attempt to look at the effectiveness of policies, no such exercise has been conducted in Flanders: studies are limited to the analysis of outputs and goal attainment without going further to establish causal connections. Hence, no causal models, whether in the form of multiple causality, theories of change or counterfactuals (cf. supra) have been developed to conceptualize the exact relationship between policies and their effects. Given the fact that the effectiveness criterion is one of the pillars of Flemish civic integration policies (cf. supra, introduction), it is evident that causality should occupy an essential role in a Flemish evaluation framework.
4. **Limited thematic focus on civic integration.** Flemish research has focused so far on the (compulsory) civic integration policies, which are composed of social

orientation courses, Dutch as a second language and career orientation. Other policy measures which have already been evaluated in other countries, such as mentoring programs, have not been assessed in Flanders due to the fact that they have recently been started. In other words, we need an evaluation framework that can be applied to both the civic integration policies that have been the focus of evaluations so far, and to other measures aiming at the integration of populations of foreign origin.

Drawing on these lessons, the following section attempts to respond to the gaps that have been identified here by presenting an evaluation framework for the Flemish integration policies.

3. THE CONSTRUCTION OF A FLEMISH EVALUATION FRAMEWORK

3.1 Introduction

The previous chapter identified several gaps in the evaluations of the Flemish integration and civic integration policies that have been conducted to date: there is no common evaluation framework for the different studies; no conceptual reflection on policy processes and evaluation criteria, no attention to causality and a limited thematic focus on civic integration as opposed to integration policies as a whole.

Those gaps constitute a significant obstacle for the construction of an evidence base for Flemish civic integration policies: the lack of a common framework or discourse precludes or encumbers dialogue within the scientific field, and the lack of attention to causality and the limited focus makes it impossible to determine whether policies have been successful or not. Keeping those gaps in mind, this chapter proposes an evaluation framework for the Flemish civic integration policies.

The framework²¹ is built on the basis of the insights we obtained from the existing evaluation infrastructure in other countries. We focused specifically on Canada and Denmark, as they have a consolidated evaluation infrastructure. On the one hand, Canada has established an Evaluation Department within the ministry responsible for immigrant integration. On the other hand, the Danish Ministry of Integration has set up a Theory of Change that guides its policy process as a whole.

Drawing on the Danish and Canadian examples, we introduce our evaluation framework in three steps:

1. A conceptual reflection in which we define the purposes of the Flemish framework for the evaluation of integration policies, establish the purposes of the evaluation and delimit its scope to integration and civic integration policies.
2. The selection and conceptualization of five evaluation criteria: effectiveness, efficiency, relevance, coherence and consistency. This conceptualization constitutes a first step towards shaping a common discourse for a science-based discussion on the success of integration policies.

²¹ It should be noted that our usage of the term “framework” differs from the Canadian interpretation provided above. The Flemish evaluation “framework” is equivalent to the Canadian “evaluation policy”, as it is applied to integration policies as a whole.

3. The development of a vision on causality. As we choose to view causality as generative, the focus lies not only on *whether* policies are successful but also on *how* they work. In order to conceptualize the *how* of policies, we introduce a theory of change that illustrates the mechanisms along which integration policies produce certain outcomes. Such a vision makes it possible to contribute to evidence-based policy making and to improve policy processes in a targeted manner.

3.2 Conceptual reflection and scope

This section tackles two gaps from the current Flemish evaluations of civic integration policies: the lack of a conceptual reflection and the limited scope of the evaluations. On the one hand, it provides a definition for the evaluation of civic integration policies and sets out the purposes of that evaluation. On the other hand, it broadens the scope of the evaluations towards integration policies as a whole instead of civic integration policies, and towards the receiving society along with new- and oldcomers. By doing so, it takes into account the definition of the integration concept as a two-way process.

The benefits of such a conceptual reflection are twofold:

1. It shapes a consensus regarding what an evaluation should contain, and what should be evaluated.
2. It provides a clear focus to the evaluation agenda.

3.2.1 Definition

The evaluation of the Flemish civic integration and integration policies consists of an objective assessment of the merit or quality of one or more particular instruments belonging to those policies, based on a list of criteria and using a methodological framework adapted to the nature of the specific instruments to be evaluated.

3.2.2 Purposes of the evaluation

The evaluation of Flemish integration policies has the following purposes:

1. To fulfill the accountability functions of the Flemish evaluation policies. The evaluation of Flemish civic integration policies is mandated in Article 26 of the Civic Integration decree.²²
2. To contribute to evidence-based policy making. In view of the European Common Basic Principles on Integration, Flemish evaluations should provide empirical evidence to improve the identification of policy needs, the definition of policy goals, and the design and implementation of policy interventions.
3. To provide a basis for policy-learning. Flemish evaluations should provide a solid empirical basis for the exchange of information across different local, provincial, regional and national settings.

²² This requirement will be extended to the integration policies as a whole when the new decree enters into force (most likely 2013).

4. To close the current gaps in Flemish research. The evaluation of Flemish integration policies should shed light on the causal relationships between policies and their results, and on the variety of integration policy instruments other than civic integration²³.

3.2.3 Scope of the evaluation

The scope of the evaluation is domain-specific and applies to the functional policy domain of integration. It should further be interpreted as follows:

1. This evaluation framework applies to both the Flemish integration and civic integration policies as defined by the civic integration decree of 2003 and the integration and civic integration decree of 2012²⁴.
2. The Flemish legal definitions of integration and civic integration emphasize the nature of integration as a process involving both newcomers and the receiving society. Therefore, the evaluation of those policies should assess the evaluation criteria for both groups.

²³ It should be noted, however, that there is currently no overview of all policy instruments belonging to the integration domain. This makes it impossible to define some of the criteria such as effectiveness, efficiency and policy relevance. Hence, we refer to integration policies when possible and otherwise focus on civic integration policies.

²⁴ A possible hindrance to this focus is the fact that there is no overview of the Flemish integration policies outside the domain of civic integration.

3.3 Evaluation criteria

3.3.1 Selection

The two first criteria that should be taken into consideration by a Flemish evaluation framework are pretty much straightforward given the emphasis that policy makers have laid on them: efficiency and effectiveness. In what follows we assess the suitability of the other criteria:

Relevance. The importance of policy relevance as a criterion is considerable: it allows us to see whether policies should be there on the first place. Besides, given the constantly changing nature of migration flows, the *raison d'être* of integration policies cannot be taken for granted anymore. The fact that work on policy relevance has already been conducted (cf. supra, examples on the relevance of language courses and on the social impact of civic integration policies) provides us with a considerable advantage to operationalize the criterion.

Goal attainment. The criterion is not included in the evaluation framework, given the fact that it is very similar to the gross effectiveness of a policy (cf. supra).

Coherence. We decide to include this criterion in the evaluation framework due to the fact that there is a multiplicity of possibly overlapping instruments (social translators and interpreters, mentoring initiatives, language coaches, language policies) that may affect each other, but there is no overview, either of the measures themselves or of their interaction. In other words, in order to be able to assess the effectiveness of civic integration and integration policies, we need to disentangle their coherence first.

Consistency. Consistency is also included in the Flemish evaluation framework in function of the complex policy setting in which integration policies are embedded: all government layers (local, provincial, regional) are involved in the delivery of integration and civic integration policies. In this setting, it is necessary to have an overview of the way in which the different levels' policy priorities fit together. Moreover, this criterion is likely to acquire importance in the future in the framework of the sector's reform, which will transfer more capacities to local governments.

Sustainability. We decide to exclude this criterion due the fact that it refers to the financial sustainability of policies, which falls out of the scope of this evaluation, in which the focus lies in the contents of the policy rather than its funding.

3.3.2 Conceptualization

In order to conceptualize the evaluation criteria, we first need a definition of the components of the policy process²⁵. In what follows, we conceptualize both elements. A schematic representation of our framework is provided in the figure below.

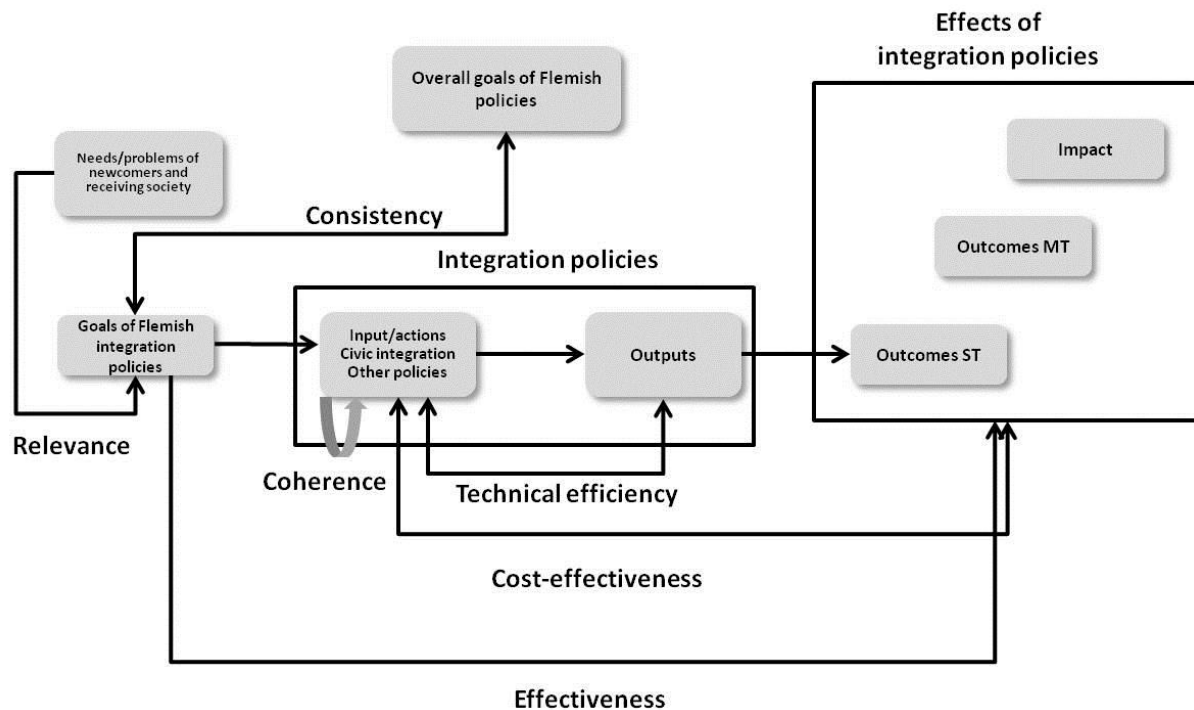


Figure 4. An evaluation framework for the Flemish integration policies

Policy needs	Policy needs are the situations, conditions of problems that trigger policies. Such a need is, for instance, the lack of knowledge of the Flemish health system by newcomers, which may trigger the inclusion of information about healthcare in Flanders in social orientation courses.
Policy goals	Policy goals are the product of or the answer to policy needs, and form the guidelines of policy interventions. Examples of goals in civic integration policies are self-sufficiency and social cohesion.
Policy inputs	Policy inputs are the means invested to realize the policy goals: money, expertise, human capital, etc.
Policy outputs	An output is a “tangible product (including services) of an intervention that is directly attributable to the initiative. Outputs relate to the completion (rather than the conduct) of activities and are the type of results over which managers

²⁵ The definition of policy needs, policy goals and policy inputs are those of Wauters. For outputs and outcomes we adopt the definition of the UNDP’s evaluation framework.

	have most influence". An example of an output for civic integration policies is a completed language course.
Policy outcomes	Outcomes are "actual or intended changes in development conditions that an intervention(s) seeks to support. The contribution of several partners is usually required to achieve an outcome". They can take place in the short-term (short-term outcomes), mid-term (mid-term outcomes) or in the long-term (impact). Examples of outcomes of civic integration policies are an enhanced language proficiency of newcomers. An example of impact is an increase in the social cohesion between the receiving society and the newcomers.
Efficiency	<p>Efficiency is the way in which policy inputs are combined in order to produce the policy results (outputs, effects). It can refer either to the relationship between inputs and outputs (cost-effectiveness), or to the way in which inputs are combined during the policy process in order to obtain a certain output (technical efficiency).</p> <p>The efficiency of civic integration policies is defined as:</p> <ul style="list-style-type: none"> a) Cost-effectiveness: the relationship between policy inputs (the monetary cost of language courses, career orientation, social orientation) and the effects. In other words, efficiency asks the question how much does it cost to make a newcomer self-sufficient, language proficient and able to design his own career path?), and judges whether the cost is reasonable with regard to the policy outcomes. b) Technical efficiency: the relationship between the deployed inputs (money, staff, knowledge, organizational processes) and the achieved outputs. When looking at this relationship, we look at the way in which inputs and outputs are related throughout the implementation process. This relationship is expressed in the following questions: How much does a civic integration certificate, a finished language course and a finished career orientation course cost? How can we deploy a given level of inputs (money, staff, knowledge, organizational processes) so that we maximize the quantity and quality of the produced outputs?
Effectiveness	<p>Effectiveness is the extent to which policy interventions lead to policy outcomes.</p> <p><u>Example:</u> Effectiveness can be assessed both for civic integration policies as a whole and for each of its components:</p> <ul style="list-style-type: none"> a) Effectiveness for civic integration policies as a whole is the extent to which the combination of policy instruments (language courses, social orientation and career orientation) lead to self-sufficiency, equal participation of the target group, and active and shared citizenship for everyone, as well as to the achievement of social cohesion. b) Effectiveness for each of the instruments is defined as follows: <ul style="list-style-type: none"> a. The effectiveness of social orientation is the extent to which following a course leads the newcomer to an enhanced knowledge of his rights and obligations, insights in our society and its basic values and the development of the competences needed for self-sufficiency.

	<ul style="list-style-type: none"> b. The effectiveness of the Dutch language courses is the extent to which following a course leads to a basic Dutch language proficiency that forms a basis to a follow-up course. c. The effectiveness of career orientation is the extent to which the newcomer acquires insight in the labor market, the education system and social activities, and sets out his own career path.
Relevance	<p>Relevance is the extent to which policy goals meet policy needs. Within the scope of this evaluation framework, we interpret relevance in terms of:</p> <ul style="list-style-type: none"> a) The fit between policy goals and needs. The assessment of those needs includes both newcomers and the receiving society. b) The uniqueness of the measure. In other words: are there any similar programs that meet the same goals?
Coherence	<p>Coherence is the internal cohesion of a policy initiative within a certain policy field. Within the evaluation of evaluation policies, the concept refers to the extent to which the different instruments relate to each other. Examples of such relationships are the way in which the three instruments of civic integration policies (social orientation, language courses and career orientation) reinforce each others' goals and outcomes, and the way in which those three instruments relate to the broader set of integration policy instruments such as mentoring schemes.</p>
Consistency	<p>Consistency is the extent to which the goals of integration and civic integration policies fit within over-arching policy goals and goals of other policy domains. General policy goals are contained in the Flemish government agreement for the current legislative term. Other policy goals are contained in Policy Briefs and Policy memorandums. The relevant policy domains for which consistency should be checked depend on the instrument being evaluated. Examples are , work and social assistance, in which some instances such as the Public Employment Services (VDAB) and the social assistance agencies (OCMW) sometimes present the target group with contradictory requirements.</p>

3.4 View on causality

One of the main lessons we drew from the overview of Flemish evaluation studies within the field of civic integration was the lack of a vision on causality: there has been no reflection about the way in which policies are supposed to produce certain results. This section makes a choice between the different conceptions of causality, based on three criteria related to the features of Flemish integration policies and the availability of information regarding integration policies. On the basis of that conception of causality, it proposes a theory of change in which the process by which policies lead to their outcomes is set out.

3.4.1 Criteria

The effectiveness of Flemish integration and civic integration policies should be based on a view of causality that takes into account the following elements:

1. **The importance of the context.** The integration and civic integration policy instruments – social orientation, language courses, career orientation, social translation, etc. – are embedded in a social, economic and cultural context that exposes newcomers to several stimuli such as daily interactions with the local population or economic phenomena. In other words, we have no control of the newcomer's environment. Those stimuli, as well as the possible interactions between overlapping instruments, need to be taken into account when evaluating integration policies.
2. **The importance of policy processes to stimulate policy learning.** Integration and civic integration policies in Flanders are locally embedded: the organization of civic integration programs may vary across welcome offices. This makes the policy's effectiveness dependent on processes taking place in that local context. In order to make policy learning possible, we need to know the causal mechanisms behind those variations in effectiveness.
3. **The availability of sources.** The available information is both quantitative (administrative data) and qualitative (information from policy actors). The quantitative sources make it possible to set up quasi-experiments, whereas the qualitative sources make it possible to trace back policy processes and to take the policy context into account.
4. **The effectiveness ladder.** We recognize that we cannot have full certainty about the effectiveness of integration and civic integration policies, because they are not set up as Randomized Control Trials. Therefore, we assess effectiveness in terms

of the effectiveness ladder. In other words, we analyze the potential, plausible and functional effectiveness of the policy instruments.

Hence, the evaluation of Flemish integration and civic integration policies regards the relationship between policy actions and policy effects in terms of:

- Counterfactuals, which allow the comparison across local cases.
- Generative causality, which allows us to incorporate the policy's context by means of a theory of change.
- Potential, plausible and functional effectiveness

3.4.2 Theory of change

As it has become clear from the previous sections, the choice for a generative causality implies that we look at *how* policies produce certain outcomes. This form of causality implies, in other words, that a lot of attention is paid to policy processes in order to facilitate the steering of policies. In order to map those processes we use *theories of change* as the main tool, as they allow us to map the processes and test that mapping in a systematic way while taking all possible determinants of an outcome into account.

i. Definition

A theory of change is the depiction of a policy theory. That theory explains how a policy intervention leads to policy results, and is composed of two elements (Chen, 2006):

1. A **description of the policy intervention** and of the way in which it is supposed to reach certain goals or effects.
2. A **view about what needs to be done** to implement policies in function of the desired results.

According to Chen, the theory of change is composed of:

1. **Policy intervention:** a description of the policy actions.
2. **Determinants²⁶:** the consequences of policy interventions at the level of the newcomer or the receiving society.
3. **Policy results:** the final effects of the policy, which should be the same as the policy goals.
4. **Moderating variables:** the elements of the context, such as the demographic features of the target group, that function as scope conditions and influence the causal relationships between policy interventions, determinants and results. For instance, a certain policy may work for an "average" newcomer, but not for an illiterate one.

²⁶ An alternative term is "intervening variables".

Figure 5 contains an example of a theory of change²⁷.

²⁷ The moderating variables are not described in detail for the sake of visual clarity. In the context of civic integration policies, those variables may be related to characteristics of the population or the external context.

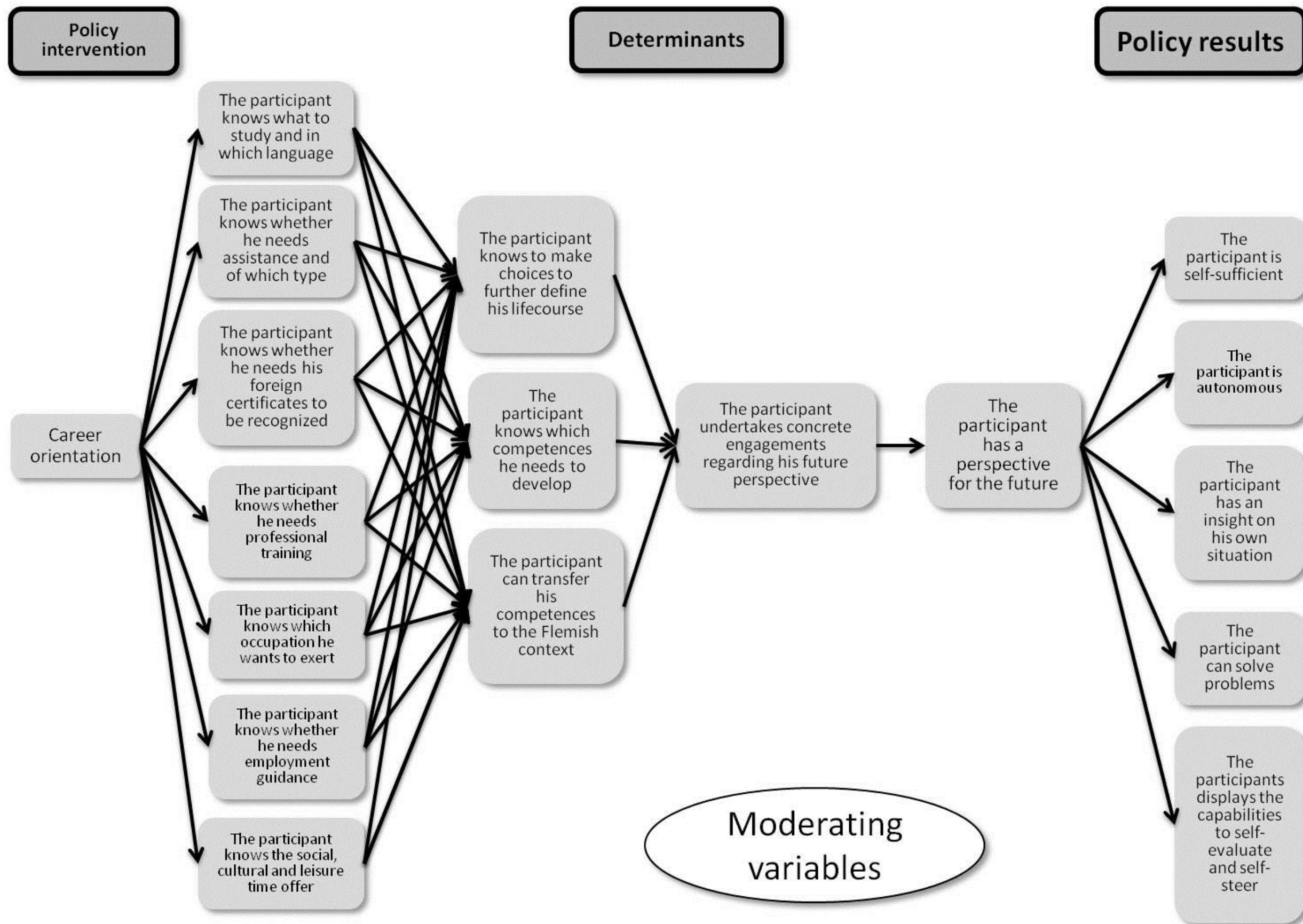


Figure 5. A theory of change for the primary pathways of the Flemish civic integration policies

Even though the evaluation framework provided above focuses on both the Flemish integration *and* civic integration policies, the theory of change cannot include *all* of the existing instruments due to three reasons:

1. First, there is no full overview of all integration policies.
2. Second, given the large number of instruments we cannot evaluate all of them at the same time. Hence, we focus the theory of change on the primary pathway of the civic integration instruments. Hence, “integration” and the various modalities of equal participation (in work, education, etc) do not appear in the theory of change, as they are constitute the long-term impact of the civic integration program *as a whole* (i.e. primary and secondary pathways).
3. Separate theories of change need to be developed for other integration instruments.

4. CONCLUSION

Drawing from a review of theoretical and policy literature on evaluation and evaluation frameworks, this paper formulates an evaluation framework to objectively assess the quality or merit of policy instruments.

That contribution is key in the current context of policy learning, both at the Flemish and the international level. At the Flemish level, it:

1. **Suggests three new criteria** to evaluate Flemish integration and civic integration policies.
2. **Provides a definition for efficiency and effectiveness**, which occupy a central position in policy processes but which remained undefined so far.
3. **Identifies the current gaps in the Flemish policy research** on the evaluation of integration and civic integration policies.
4. **Establishes a basis for comparing local integration practices**, which are likely to gain importance in the framework of the current reform of the integration sector.
5. **Lays the ground for the construction of an evaluation framework** that allows to evaluate Flemish policies on a systematic basis.

At the international level, the evaluation framework provides a tool to enhance the evidence-based contents of the Flemish policy experience. This provides us with a lever to increase the visibility of Flemish policies at the EU and OECD level, and responds to the EU's CBP that stresses the development of "clear goals, indicators and evaluation mechanisms".

The evaluation framework should, however, only be regarded as the first step in the long process of constructing an infrastructure for evidence-based policy-making. In this sense, we can identify several lines of action that should be followed:

1. **The evaluation framework needs to be further embedded in an official policy document**, and disseminated as the blueprint for future evaluations. Besides defining the evaluation criteria and the scope of the evaluation, such a document should provide concrete guidelines for carrying out evaluations, such as detailed instructions about the periodicity of evaluations, whether the evaluation should be internal or external and a deontological code.
2. **The mapping of existing integration instruments to define the framework's scope**. The framework's scope is domain-specific: it covers both

civic integration policies and the broader spectrum of integration policies. However, those instruments have not been mapped so far. Therefore it is necessary, before any evaluations are conducted, to have an exhaustive overview of the existing instruments and data sources regarding those instruments.

3. **The operationalization of the evaluation framework.** Whereas the framework provides a solid structure on which to base future evaluations, it remains at a generic level and needs to be further specified in the same way as the CIC policy is further developed in instrument-specific studies. There are several elements that can only be further refined by individual evaluation studies, starting with the operationalization of the policy process. Policy goals such as social cohesion, self-sufficiency and participation are complex constructs that should be properly conceptualized and measured.
4. **The coordination of evaluation with existing and new monitoring mechanisms.** In order to achieve a sound operationalization, the evaluation framework needs to be linked to the monitoring infrastructure of civic integration and integration policies (as in the case of South Africa, where a monitoring and evaluation framework has been developed). By mapping out the existing monitoring sources we can obtain some insights in what can be achieved by evaluations.
5. **The development of methodological strategies.** Such strategies should focus on the collection of data complementary to that obtained from monitoring sources, and on the development of techniques. Foreign studies, such as the Danish ones, can serve as an inspiration in the process.

The policies that constitute the object of the evaluation framework are at the moment a moving target: the integration sector is being subjected to a structural reform in which service provision will undergo several changes, such as the disappearance of career orientation as a stand-alone instrument of civic integration policies. It is clear, however, that evaluation will remain a priority for the coming years. In this sense, this evaluation framework will prove an essential instrument to keep track of the policy changes and the way in which those changes affect policy performance.

REFERENCES

- Auer, P., & Kruppe, T. (1996). *Monitoring of Labour Market Policy in EU Member States*. Berlin.
- Bonin, H., Roberts, R., & Zimmerman, K. F. (2008). Comparing and evaluating public expenditure on migration. In S. Ardittis & F. Laczko (Eds.), *Assessing the costs and impacts of migration policy: an international comparison*. Geneva and Cheshire: IOM and Eurasyllum.
- Caviedes, A. (2004). The open method of co-ordination in immigration policy: a tool for prying open Fortress Europe? *Journal of European Public Policy*, 11(2), 289–310.
- Chen, H. T. (2006). A theory-driven evaluation perspective on mixed methods research. *Research in the Schools*, 13(1), 75–83.
- Citizenship and Immigration Canada. (2012). CIC Evaluation Policy. Retrieved November 12, 2012, from <http://www.cic.gc.ca/english/resources/evaluation/policy.asp>
- De Cuyper, P. (2010). *Deel 2. Het inburgeringsbeleid in Vlaanderen: de efficiëntie en effectiviteit geëvalueerd*. Leuven: HIVA.
- De Cuyper, P., De Rick, K., & Gonzalez Garibay, M. (2012). *Persoonlijke ontwikkelingsplannen: concept en implementatie*. Leuven: HIVA.
- De Cuyper, P., & Jacobs, L. (2011). *Het NT2-aanbod in Vlaanderen: passend voor werkenden en werkzoekenden?* Leuven: HIVA.
- De Cuyper, P., Lamberts, M., & Pauwels, F. (2010). *De efficiëntie, effectiviteit en impact van het Vlaamse inburgeringsbeleid geëvalueerd. Synthese*. Leuven: HIVA.
- De Cuyper, P., & Wets, J. (2007). *Diversiteit in integratie. Een evaluatie van de vormgeving, efficiëntie en effectiviteit van het Vlaamse inburgeringsbeleid*. Leuven: HIVA.
- De Peuter, B., De Smedt, J., & Bouckaert, G. (2007). *Handleiding beleidsevaluatie. Deel 1: Evaluatiedesign en -management*. Leuven: SBOV.
- Delander, L., Hammarstedt, M., Månsson, J., & Nyberg, E. (2005). Integration of immigrants: the role of language proficiency and experience. *Evaluation review*, 29(1), 24–41.
- Development Assistance Committee. (n.d.). *Principles for evaluation of development assistance*. Paris: OECD.
- Eurostat. (2011). *Indicators of immigrant integration. A pilot study*. Brussels.
- Evaluation Division. (2010a). *Evaluation of the Host Program*. Ottawa: Citizenship and Immigration Canada.

- Evaluation Division. (2010b). *Evaluation of the Language Instruction for Newcomers to Canada (LINC) Program*. Ottawa: Citizenship and Immigration Canada.
- Evaluation Division. (2012). *Evaluation of the Multiculturalism Program*. Ottawa: Citizenship and Immigration Canada.
- Fournier, D. M. (2005). Evaluation. In S. Mathison (Ed.), *Encyclopedia of evaluation* (pp. 139–140). Thousand Oaks: SAGE.
- Gysen, J. (2006). *Geen vuiltje aan de bodem? Een evaluatie van de effectiviteit en de neveneffecten van het bodemsaneringsdecreet*. Leuven: HIVA.
- Health Service Executive. (2008). *National Intercultural Health Strategy 2007-2012*. Dublin: Health Service Executive.
- Healy, G. (2010). *ICI Mentoring Programme : an independent evaluation. Final report*.
- Horwath Consulting Ireland, Ramboll Management, & Matrix Knowledge Group. (2008). *Final Report: Office of the Minister for Integration and the Department of Education & Science. Development of a National English Language Policy and Framework for Legally-Resident Adult Immigrants*.
- Howlett, M., Ramesh, M., & Perl, A. (2009). *Studying public policy. Policy cycles & policy subsystems* (Third.). Don Mills, Ontario: Oxford University Press.
- Humanitarian Branch DIAC. (2009). *The evaluation of AUSCO 2009*.
- IMDi, & VOX. (2010). Making a success of integrating immigrants into the labour market. Host country report. Retrieved November 12, 2012, from <http://www.peer-review-social-inclusion.eu/peer-reviews/2010/making-a-success-of-integrating-immigrants-into-the-labour-market>
- IOM. (2010). *Avaliação de acompanhamento. Sistema Nacional de Apoio ao Imigrante*.
- Joona, P. A., & Nekby, L. (2012). Intensive Coaching of New Immigrants: An Evaluation Based on Random Program Assignment*. *The Scandinavian Journal of Economics*, 114(2), 575–600.
- Luedtke, A. (2009). Uncovering European Union Immigration Legislation : Policy Dynamics and Outcomes. *International Migration*, 49(2).
- Mayne, J. (2011). Contribution analysis: addressing cause and effect. In K. Forrs, M. Marra, & R. Schwarz (Eds.), *Evaluating the complex. Attribution, contribution and beyond* (pp. 53–96). New Brunswick, NJ: Transaction Publishers.
- Moller Hansen, L. (2012). Monitoring the integration process in Denmark. In R. Bijl & A. Verweij (Eds.), *Measuring and monitoring immigrant integration in Europe*. The Hague: The Netherlands Institute for Social Research.
- OECD. (2007). *Jobs for Immigrants. Volume 1. Labour market integration Australia, Denmark, Germany and Sweden* (Vol. 1). Paris: OECD Publishing.

- OECD. (2008a). *The Paris declaration on aid effectiveness and the Accra agenda for action*. Paris: OECD.
- OECD. (2008b). *Jobs for Immigrants. Volume 2. Labour market integration in Belgium, France, the Netherlands and Portugal. 2008*. Paris: OECD Publishing.
- OECD. (2012). *Jobs for Immigrants. Volume 3. Labour market integration in Austria, Norway and Switzerland (Vol. 3)*. Paris: OECD Publishing.
- Owen, J. M., & Rogers, P. J. (1999). *Program evaluation. Forms and approaches*. Thousand Oaks: SAGE.
- Pauwels, F., & Lamberts, M. (2010). *Deel 3 . De impact van inburgering*. Leuven: HIVA.
- Rossi, P. H., Lipsey, M. W., & Freeman, H. E. (2004). *Evaluation. A systematic approach*. Thousand Oaks: SAGE.
- Royse, D., Thyer, B. A., Padgett, D. K., & Logan, T. (2006). *Program evaluation. An introduction*. Belmont, CA: Thomson Brooks/Cole.
- Russon, C. (2005). Cluster evaluation. In S. Mathison (Ed.), *Encyclopedia of evaluation* (pp. 66–67). Thousand Oaks: SAGE.
- Schibel, Y., Fazel, M., Robb, R., & Garner, P. (2002). Refugee integration: Can research synthesis inform policy? Feasibility study report. *RDS On-line Report 13/02*. Retrieved April 11, 2013, from <http://library.npia.police.uk/docs/hordsolr/rdsolr1302.pdf>
- Schuller, K., Lochner, S., & Rother, N. (2011). *Integration Panel. Results of a longitudinal study on the effectiveness and sustainability of integration courses*. (pp. 1–13). Federal Office for Migration and Refugees.
- Shaw, I. F., Greene, J. C., & Mark, M. M. (Eds.). (2006). *The SAGE handbook of evaluation*. Thousand Oaks: SAGE.
- Significant. (2010). *Evaluatierapport inburgering in Nederland*. Barneveld: Significant.
- Stern, E., Stame, N., Mayne, J., Forss, K., Davies, R., & Befani, B. (2012). Broadening the range of designs and methods for impact evaluations.
- Swanborn, P. G. (1999). *Evalueren*. Amsterdam: Boom.
- The Presidency. (2007). *Policy framework for the government-wide monitoring and evaluation system*. Pretoria.
- Treasury Board of Canada Secretariat. (2012). Policy on Evaluation. Retrieved November 12, 2012, from <http://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=15024§ion=text>
- Urbis. (2003). *Evaluation of the Integrated Humanitarian Settlement Strategy (IHSS)*. Belconnen: Urbis.

- Veerman, J. W., & Van Yperen, T. A. (2007). Degrees of freedom and degrees of certainty: a developmental model for the establishment of evidence-based youth care. *Evaluation and Program Planning*, 30(2), 212–221.
- Verweij, A. (2012). Policy concepts in the Netherlands: integration, super-diversity and social trust. Retrieved November 12, 2012, from <http://www.steunpuntieni.be/main.aspx?c=.SII&n=110202>
- W.K. Kellogg Foundation. (2004). *Evaluation Handbook*. Battle Creek, MI: W.K. Kellogg Foundation.
- Wets, J., Seghers, M., Pauwels, F., De Cuyper, P., & Van Avermaet, P. (2012). *Het civiel effect van inburgering: conceptuele beschouwingen en nulmeting*. Leuven: HIVA.
- Wholey, J. S., Hatry, H. P., & Newcomer, K. E. (Eds.). (1994). *Handbook of practical program evaluation*. San Francisco: Jossey-Bass.
- Yanasmayan, Z., & Foblets, M.-C. (2012). Pros and cons of compulsory “inburgering”: Mapping the views of newcomers in Flanders and in Brussels. *Which Integration Policies for Migrants? Interaction between the EU and its Member States* (pp. 255–285). Oisterwijk: Wolf Legal Publishers.



Steunpunt Inburgering en Integratie

Bezoekadres: Lange Nieuwstraat 55

Postadres: Prinsstraat 13

B-2000 Antwerpen

Tel.: +32 3 265 59 63

E-mail: steunpuntieni@ua.ac.be

Website: <http://www.steunpuntieni.be>

Contact auteurs:

Montserrat González Garibay,

montserrat.gonzalezgaribay@kuleuven.be - +32 16 323318

Peter De Cuyper,

peter.decuyper@kuleuven.be +32 16 323175